(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 June 2001 (21.06.2001)

PCT

(10) International Publication Number
WO 01/44445 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/US00/33815

(22) International Filing Date:
13 December 2000 (13.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/172,307 16 December 1999 (16.12.1999) US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Moutain View, CA 94043 (US). **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive #17, Mountain View, CA 94040 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95129 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/44445 A2

(54) Title: HUMAN LYASES AND ASSOCIATED PROTEINS

(57) Abstract: The invention provides human lyases and associated proteins (HLYAP) and polynucleotides which identify and encode HLYAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of HLYAP.

HUMAN LYASES AND ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human lyases and associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of reproductive and neurological disorders, inflammatory disorders, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human lyases and associated proteins.

BACKGROUND OF THE INVENTION

Lyases are a class of enzymes that catalyze the cleavage of C-C, C-O, C-N, C-S, C-(halide), P-O, or other bonds without hydrolysis or oxidation to form two molecules, at least one of which contains a double bond (Stryer, L. (1995) Biochemistry, W.H. Freeman and Co., New York NY, p.620). Under the International Classification of Enzymes (Webb, E.C. (1992) Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes, Academic Press, San Diego CA), lyases form a distinct class designated by the numeral 4 in the first digit of the enzyme number (i.e., EC 4.x.x.x).

Further classification of lyases reflects the type of bond cleaved as well as the nature of the cleaved group. The group of C-C lyases includes carboxyl-lyases (decarboxylases), aldehyde-lyases (aldolases), oxo-acid-lyases, and other lyases. The C-O lyase group includes hydro-lyases, lyases acting on polysaccharides, and other lyases. The C-N lyase group includes ammonia-lyases, amidine-lyases, amine-lyases (deaminases), and other lyases.

Lyases are critical components of cellular biochemistry, with roles in metabolic energy production, including fatty acid metabolism and the tricarboxylic acid cycle, as well as other diverse enzymatic processes.

Phosphoenolpyruvate carboxykinase (ATP) (EC 4.1.1.49) is a lyase involved in gluconeogenesis, the production of glucose from storage compounds in the body. This enzyme catalyzes the decarboxylation of oxaloacetate to form phosphoenolpyruvate, accompanied by hydrolysis of ATP. (See, e.g., Matte, A. et al. (1997) J. Biol. Chem. 272:8105-8108; Medina, V. et al. (1990) J. Bacteriol. 172:7151-7156.)

L-rhamnose and D-fucose are 6-deoxyhexoses found in complex carbohydrates in bacterial cell walls. One of the steps in the pathways leading to the synthesis of these carbohydrates is the conversion of dTDP-D-glucose to an unstable 4-keto-6-deoxy intermediate, a reaction catalyzed by the lyase dTDP-D-glucose 4,6-dehydratase (EC 4.2.1.46). (See, e.g., Tonetti, M. et al. (1998)

Biochimie 80:923-931; Yoshida, Y. et al. (1999) J. Biol. Chem. 274:16933-16939.)

Isocitrate lyase (EC 4.1.3.1) is involved in the glyoxylate cycle, a modification of the citric acid cycle. The glyoxylate cycle occurs in bacteria, fungi, and plants. Isocitrate lyase catalyzes the cleavage of isocitrate to yield succinate and glyoxylate. (See, e.g., Beeching, J.R. (1989) Protein Seq. Data Anal. 2:463-466; Atomi, H. et al. (1990) J. Biochem. 107:262-266.)

Aldolases are lyases which catalyze aldol condensation reactions. Fructose 1,6-bisphosphate aldolase (FBP-aldolase; EC 4.1.2.13) catalyzes the reversible cleavage of fructose 1,6-bisphosphate to yield dihydroxyacetone phosphate, a ketose, and glyceraldehyde 3-phosphate, an aldose. Class I FBP-aldolases are found in higher organisms, and exist as homotetramers. Class II FBP-aldolases tend to be dimeric, occur in yeast and bacteria, and have an absolute requirement for a divalent cation for catalytic activity. (See, e.g., Hall, D.R. et al. (1999) J. Mol. Biol. 287:383-394.)

Pseudouridine is an isomer of uridine which helps to maintain the specific tertiary structures of certain rRNAs, tRNAs, and small nuclear and nucleolar RNAs. Pseudouridine is not directly incorporated into these RNAs, but is synthesized by pseudouridine synthases (EC 4.2.1.70), lyases which act on specific uridine residues within these RNAs. The Rlu family of pseudouridine synthases includes Escherichia coli ribosomal large subunit synthase A, which synthesizes pseudouridine at position 746 in 23S rRNA and Escherichia coli ribosomal large subunit synthase C, which synthesizes pseudouridine at positions 955, 2504, and 2580 in 23S rRNA. (See, e.g., Conrad, J. et al. (1998) J. Biol. Chem. 273:18562-18566.)

Fumarate lyases are a group of lyases which share limited sequence homology and use fumarate as a substrate. These enzymes include fumarase (EC 4.2.1.2), aspartase (EC 4.3.1.1), arginosuccinase (EC 4.3.2.2), and adenylosuccinase (EC 4.3.2.2). (See, e.g., Woods, S.A. et al. (1988) Biochim. Biophys. Acta 954:14-26; Woods, S.A. et al. (1988) FEMS Microbiol. Lett. 51:181-186; Zalkin, H. and J.E. Dixon (1992) Prog. Nucleic Acid Res. Mol. Biol. 42:259-287.)

The glyoxalase system is involved in gluconeogenesis, the production of glucose from storage compounds in the body. It consists of the lyase glyoxalase I (EC 4.4.1.5), which catalyzes the formation of S-D-lactoylglutathione from methylglyoxal, a side product of triose-phosphate energy metabolism, and glyoxalase II, which hydrolyzes S-D-lactoylglutathione to D-lactic acid and reduced glutathione. Glyoxalases are involved in hyperglycemia, non-insulin-dependent diabetes mellitus, the detoxification of bacterial toxins, and the control of cell proliferation and microtubule assembly. (See, e.g., Thornalley, P.J. (1993) Mol. Aspects Med. 14:287-371.)

Aconitase (EC 4.2.1.3) is a lyase which carries out a crucial step in the tricarboxylic acid cycle. Aconitase catalyzes the reversible transformation of citrate into isocitrate through a cis-aconitate intermediate. Two forms of aconitase are found in mammalian cells, a cytosolic aconitase (Kennedy, M.C. et al. (1992) Proc. Natl. Acad. Sci. USA 89:11730-11734) and a mitochondrial

aconitase (Mirel, D.B. et al. (1998) Gene 213:205-218).

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco; EC 4.1.1.39) is a lyase which carries out a crucial step in the Calvin cycle during photosynthesis. Rubisco catalyzes the covalent incorporation of carbon dioxide into the 5-carbon sugar ribulose 1,5-bisphosphate along with the simultaneous cleavage of this molecule into two molecules of 3-phosphoglycerate. (See, e.g., Hartman, F.C. and M.R. Harpel (1994) Annu. Rev. Biochem. 63:197-234.) Specific methyltransferases (EC 2.1.1.43) catalyze the methylation of amino groups near the N-termini of the small and large subunits of Rubisco (Ying, Z. et al. (1998) Acta Biol. Hung. 49:173-184; Klein, R.R. and R.L. Houtz (1995) Plant Mol. Biol. 27:249-261).

Proper regulation of lyases is critical to normal physiology. For example, mutation induced deficiencies in the uroporphyrinogen decarboxylase can lead to photosensitive cutaneous lesions in the genetically-linked disorder familial porphyria cutanea tarda (Mendez, M. et al. (1998) Am. J. Genet. 63:1363-1375). It has also been shown that adenosine deaminase (ADA) deficiency stems from genetic mutations in the ADA gene, resulting in the disorder severe combined immunodeficiency disease (SCID) (Hershfield, M.S. (1998) Semin. Hematol. 35:291-298).

The discovery of new human lyases and associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of reproductive and neurological disorders, inflammatory disorders, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human lyases and associated proteins.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, human lyases and associated proteins, referred to collectively as "HLYAP" and individually as "HLYAP-1," "HLYAP-2," "HLYAP-3," "HLYAP-4," "HLYAP-5," "HLYAP-6," "HLYAP-7," "HLYAP-8," "HLYAP-9," and "HLYAP-10." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an

amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-10. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:11-20.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid

sequence selected from the group consisting of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence

having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional HLYAP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional HLYAP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional HLYAP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds

to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:11-20, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii),

and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding HLYAP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of HLYAP.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding HLYAP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"HLYAP" refers to the amino acid sequences of substantially purified HLYAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of HLYAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of HLYAP either by directly interacting with HLYAP or by acting on components of the biological pathway in which HLYAP participates.

An "allelic variant" is an alternative form of the gene encoding HLYAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HLYAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as HLYAP or a polypeptide with at least one functional characteristic of HLYAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HLYAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HLYAP.

The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent HLYAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of HLYAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of HLYAP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of HLYAP either by directly interacting with HLYAP or by acting on components of the biological pathway in which HLYAP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind HLYAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which

bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic HLYAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding HLYAP or fragments of HLYAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for

fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of HLYAP or the polynucleotide encoding HLYAP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:11-20 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:11-20, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:11-20 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:11-20 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:11-20 and the region of SEQ ID NO:11-20 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-10 is encoded by a fragment of SEQ ID NO:11-20. A fragment of SEQ ID NO:1-10 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-10. For example, a fragment of SEQ ID NO:1-10 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-10. The precise length of a fragment of SEQ ID NO:1-10 and the region of SEQ ID NO:1-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a

standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as

defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for

example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may

be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of HLYAP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of HLYAP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of HLYAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of HLYAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the

antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an HLYAP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of HLYAP.

"Probe" refers to nucleic acid sequences encoding HLYAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs

can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a

vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding HLYAP, or fragments thereof, or HLYAP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells,

trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook, J. et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may

possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human lyases and associated proteins (HLYAP), the polynucleotides encoding HLYAP, and the use of these compositions for the diagnosis, treatment, or prevention of reproductive and neurological disorders, inflammatory disorders, and cell proliferative disorders, including cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding HLYAP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each HLYAP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each HLYAP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding HLYAP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:11-20 and to distinguish between SEQ ID NO:11-20 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express HLYAP as a fraction of total tissues expressing HLYAP. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing HLYAP as a fraction of total tissues expressing HLYAP. Column 5 lists the vectors used to subclone each cDNA library.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding HLYAP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

SEQ ID NO:13 maps to chromosome 1 within the interval from 213.2 to 222.7 centiMorgans. SEQ ID NO:19 maps to chromosome 14 within the interval from 112.6 to 116.3 centiMorgans.

The invention also encompasses HLYAP variants. A preferred HLYAP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the HLYAP amino acid sequence, and which contains at least one functional or structural characteristic of HLYAP.

The invention also encompasses polynucleotides which encode HLYAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20, which encodes HLYAP. The polynucleotide sequences of SEQ ID NO:11-20, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding HLYAP. In particular, such a variant polynucleotide sequence will have at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding HLYAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20 which has at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-20. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HLYAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding HLYAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring HLYAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode HLYAP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring HLYAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HLYAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HLYAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode HLYAP and HLYAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HLYAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:11-20 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence

preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding HLYAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HLYAP may be cloned in recombinant DNA molecules that direct expression of HLYAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express HLYAP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HLYAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of HLYAP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of

homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding HLYAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, HLYAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of HLYAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active HLYAP, the nucleotide sequences encoding HLYAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding HLYAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HLYAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding HLYAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding HLYAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HLYAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding HLYAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding HLYAP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding HLYAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro

transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of HLYAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of HLYAP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of HLYAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, supra; and Scorer, supra.)

Plant systems may also be used for expression of HLYAP. Transcription of sequences encoding HLYAP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, supra; Broglie, supra; and Winter, supra.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding HLYAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses HLYAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of HLYAP in cell lines is preferred. For example, sequences encoding HLYAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous

expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycos.des neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HLYAP is inserted within a marker gene sequence, transformed cells containing sequences encoding HLYAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HLYAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding HLYAP and that express HLYAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of HLYAP using either

specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HLYAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HLYAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HLYAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HLYAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HLYAP may be designed to contain signal sequences which direct secretion of HLYAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing

of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding HLYAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric HLYAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of HLYAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the HLYAP encoding sequence and the heterologous protein sequence, so that HLYAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled HLYAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

HLYAP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to HLYAP. At least one and up to a plurality of test compounds may be screened for specific binding to HLYAP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of HLYAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which HLYAP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express HLYAP, either as a secreted

protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing HLYAP or cell membrane fractions which contain HLYAP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either HLYAP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with HLYAP, either in solution or affixed to a solid support, and detecting the binding of HLYAP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

HLYAP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of HLYAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for HLYAP activity, wherein HLYAP is combined with at least one test compound, and the activity of HLYAP in the presence of a test compound is compared with the activity of HLYAP in the absence of the test compound. A change in the activity of HLYAP in the presence of the test compound is indicative of a compound that modulates the activity of HLYAP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising HLYAP under conditions suitable for HLYAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of HLYAP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding HLYAP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell

blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding HLYAP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding HLYAP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding HLYAP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress HLYAP, e.g., by secreting HLYAP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of HLYAP and human lyases and associated proteins. In addition, the expression of HLYAP is closely associated with reproductive and nervous tissue, inflammation, cell proliferation, and cancer. Therefore, HLYAP appears to play a role in reproductive and neurological disorders, inflammatory disorders, and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased HLYAP expression or activity, it is desirable to decrease the expression or activity of HLYAP. In the treatment of disorders associated with decreased HLYAP expression or activity, it is desirable to increase the expression or activity of HLYAP.

Therefore, in one embodiment, HLYAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HLYAP. Examples of such disorders include, but are not limited to, a reproductive disorder, such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, a disruption of spermatogenesis, abnormal sperm

physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer,

including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing HLYAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HLYAP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified HLYAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HLYAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of HLYAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HLYAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of HLYAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HLYAP. Examples of such disorders include, but are not limited to, those reproductive and neurological disorders, inflammatory disorders, and cell proliferative disorders, including cancer, described above. In one aspect, an antibody which specifically binds HLYAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express HLYAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HLYAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HLYAP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of HLYAP may be produced using methods which are generally known in the art. In particular, purified HLYAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HLYAP. Antibodies to HLYAP may also be generated using methods that are well known in the art. Such antibodies may include, but are

not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HLYAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HLYAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of HLYAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to HLYAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HLYAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter,

G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for HLYAP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between HLYAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HLYAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for HLYAP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of HLYAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple HLYAP epitopes, represents the average affinity, or avidity, of the antibodies for HLYAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular HLYAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the HLYAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of HLYAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of HLYAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and

Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding HLYAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding HLYAP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding HLYAP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding HLYAP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399),

hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in HLYAP expression or regulation causes disease, the expression of HLYAP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in HLYAP are treated by constructing mammalian expression vectors encoding HLYAP and introducing these vectors by mechanical means into HLYAP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of HLYAP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). HLYAP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding HLYAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with

respect to HLYAP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding HLYAP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding HLYAP to cells which have one or more genetic abnormalities with respect to the expression of HLYAP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544; and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding HLYAP to target cells which have one or more genetic abnormalities with respect to the expression of HLYAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing HLYAP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with

ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding HLYAP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for HLYAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of HLYAP-coding RNAs and the synthesis of high levels of HLYAP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of HLYAP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can

be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HLYAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HLYAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding HLYAP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased HLYAP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding HLYAP may be therapeutically useful, and in the treament of disorders associated with decreased HLYAP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding HLYAP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding HLYAP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding HLYAP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding HLYAP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide

sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of HLYAP, antibodies to HLYAP, and mimetics, agonists, antagonists, or inhibitors of HLYAP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising HLYAP or fragments thereof. For example, liposome preparations

containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, HLYAP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HLYAP or fragments thereof, antibodies of HLYAP, and agonists, antagonists or inhibitors of HLYAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells,

conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind HLYAP may be used for the diagnosis of disorders characterized by expression of HLYAP, or in assays to monitor patients being treated with HLYAP or agonists, antagonists, or inhibitors of HLYAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HLYAP include methods which utilize the antibody and a label to detect HLYAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring HLYAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of HLYAP expression. Normal or standard values for HLYAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to HLYAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of HLYAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding HLYAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of HLYAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of HLYAP, and to monitor regulation of HLYAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HLYAP or closely related molecules may be used to identify nucleic acid sequences which encode HLYAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding HLYAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the HLYAP encoding sequences. The hybridization probes of the subject

invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:11-20 or from genomic sequences including promoters, enhancers, and introns of the HLYAP gene.

Means for producing specific hybridization probes for DNAs encoding HLYAP include the cloning of polynucleotide sequences encoding HLYAP or HLYAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HLYAP may be used for the diagnosis of disorders associated with expression of HLYAP. Examples of such disorders include, but are not limited to, a reproductive disorder, such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial

frontotemporal dementia; an inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer, including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding HLYAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered HLYAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HLYAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding HLYAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding HLYAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of HLYAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding HLYAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HLYAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding HLYAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding HLYAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding HLYAP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding HLYAP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal

tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of HLYAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for HLYAP, or HLYAP or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to

generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample

containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for HLYAP to quantify the levels of HLYAP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should

be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding HLYAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1

constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding HLYAP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, HLYAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between HLYAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with HLYAP, or fragments thereof, and washed. Bound HLYAP is then detected by methods well known in the art. Purified HLYAP can also be coated directly onto plates for use in the aforementioned drug screening techniques.

Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HLYAP specifically compete with a test compound for binding HLYAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HLYAP.

In additional embodiments, the nucleotide sequences which encode HLYAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/172,307, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the

recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were

carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and

amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:11-20. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \times \text{minimum } \{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding HLYAP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the

sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Chromosomal Mapping of HLYAP Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:11-20 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:11-20 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

The genetic map locations of SEQ ID NO:13 and SEQ ID NO:19 are described in The Invention as ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VI. Extension of HLYAP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:11-20 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension

was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI.B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2,

v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:11-20 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:11-20 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470;

Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5

μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60 °C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60 °C. The arrays are washed for 10 min at 45 °C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45 °C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

IX. Complementary Polynucleotides

Sequences complementary to the HLYAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring HLYAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of HLYAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HLYAP-encoding transcript.

X. Expression of HLYAP

Expression and purification of HLYAP is achieved using bacterial or virus-based expression systems. For expression of HLYAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express HLYAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of HLYAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding HLYAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, HLYAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from HLYAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified HLYAP obtained by these methods can be used directly in the assays shown in Examples XI and XV.

XI. Demonstration of HLYAP Activity

Lyase activity of HLYAP is demonstrated through a variety of specific enzyme assays. In general, HLYAP is incubated with its substrate(s) under conditions suitable for the enzymatic reaction being assayed. After a suitable period of time, the reaction is terminated, and the formation of the

product(s) are monitored spectrophotometrically, chromatographically, fluorometrically, or by some other appropriate method. Lyase activity is proportional to the amount of product(s) formed, or the rate of product formation. Some examples of specific lyase activity assays are described below.

Glyoxalase I activity of HLYAP is measured by monitoring the formation of glutathione thioester from methylglyoxal and glutathione. HLYAP is incubated with 2mM methylglyoxal and 2 mM reduced glutathione in 0.1 M sodium phosphate, pH 7.0, at 30°C. Formation of the glutathione thioester is monitored spectrophotometrically at a wavelength of 240 nm. Glyoxalase I activity of HLYAP is proportional to the rate of formation of the glutathione thioester. (See, e.g., Ridderstrom, M. et al. (1998) J. Biol. Chem. 273:21623-21628.)

dTDP-D-glucose 4,6-dehydratase activity of HLYAP is measured by monitoring the formation of dTDP-4-keto-6-deoxy-D-glucose from dTDP-D-glucose. HLYAP is incubated with 50 mM Tris-HCl, pH 7.6, 12 mM $MgCl_2$, 4 mM dTDP-D-glucose, 0.9 unit of inorganic pyrophosphatase, and 8 mM NADPH for 3 hours at 37°C. The sugar components in the mixture are coupled with 2-aminopyridine and then analyzed chromatographically using an anion-exchange column. Dehydratase activity is proportional to the amount of dTDP-4-keto-6-deoxy-D-glucose formed. (See, e.g., Yoshida, 1999, supra.)

Aconitase activity of HLYAP is measured in an assay coupled to isocitric dehydrogenase. HLYAP is incubated with isocitric dehydrogenase, NADP, and citrate, and the reduction of NADP is monitored fluorometrically. Aconitase activity is proportional to the rate of NADP reduction. (See, e.g., Costello, L.C. et al. (1997) J. Biol. Chem. 272:28875-28881; Costello, L.C. et al. (1996) Urology 48:654-659.)

XII. Functional Assays

HLYAP function is assessed by expressing the sequences encoding HLYAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the

apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of HLYAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding HLYAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding HLYAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIII. Production of HLYAP Specific Antibodies

HLYAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the HLYAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-HLYAP activity by, for example, binding the peptide or HLYAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIV. Purification of Naturally Occurring HLYAP Using Specific Antibodies

Naturally occurring or recombinant HLYAP is substantially purified by immunoaffinity chromatography using antibodies specific for HLYAP. An immunoaffinity column is constructed by covalently coupling anti-HLYAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HLYAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HLYAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HLYAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HLYAP is collected.

XV. Identification of Molecules Which Interact with HLYAP

HLYAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HLYAP, washed, and any wells with labeled HLYAP complex are assayed. Data obtained using different concentrations of HLYAP are used to calculate values for the number, affinity, and association of HLYAP with the candidate molecules.

Alternatively, molecules interacting with HLYAP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

HLYAP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	11	168714	LIVRNOT01	168714H1 (LIVRNOT01), 168714R6 (LIVRNOT01), 1297988F6 (BRSTNOT07), 2676706H1 (KIDNNOT19), 2769123H1 (COLANOT02), 3139665H1 (SMCCNOT02), 3979530H1 (LUNGTUT08), g3917881
2	12	1851727	LUNGFET03	030436H1 (THP1NOB01), 148166H1 (FIBRNGT01), 919114R1 (BRSTNOT04), 1851727F6 (LUNGFET03), 1851727H1 (LUNGFET03), 1851727X14R1 (LUNGFET03), 1984112T6 (LUNGAST01), 2519821H1 (BRAITUT21)
3	13	2095185	BRAITUT02	269794R1 (HNT2NOT01), 691073R6 (LUNGTUT02), 1867052T7 (SKINBIT01), 2095185H1 (BRAITUT02), 2095185X11B2 (BRAITUT02), 2362184R6 (LUNGFET05), 2483023H1 (SMCANOT01), 2603859F6 (LUNGTUT07), 2847019F6 (DRGLNOT01), 4290006H1 (BRABDIR01)
4	14	2342959	TESTTUT02	495043F1 (HNT2NOT01), 2342959H1 (TESTTUT02), 2396653F6 (THP1AZT01), SXAE04649V1, SXAE01464V1
5	15	2613975	THYRNOT09	2613975H1 (THYRNOT09), 2887901F6 (SINJNOT02), 2937493H1 (THYMFET02), 4979580H1 (HELATXT04), 5735306H1 (KIDCTMT01), 5985015H1 (MCLDTXT02)
6	16	2683534	SINIUCT01	190137F1 (SYNORAB01), 2683534H1 (SINIUCT01), 3703529H1 (PENCNOT07)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
7	17	2801723	PENCNOT01	436289H1 (THYRNOT01), 716906H1 (PROSTUT01), 1210169R2 (BRSTNOT02), 1210169T1 (BRSTNOT02), 1911931F6 (CONNTUT01), 2113755H1 (BRAITUT03), 2373504F6 (ISLTNOT01), 2801723F6 (PENCNOT01), 2801723H1 (PENCNOT01), 3873447H1 (HEARNOT06)
8	18	3130234	LUNGTUT12	161050F1 (ADENINB01), 621367F1 (PGANNOT01), 1384190F6 (BRAITUT08), 1397503F6 (BRAITUT08), 2376021F6 (ISLTNOT01), SAEA01676R1, SAEA01426R1, SAEA01369F1
9	19	3256118	OVARTUN01	239425R1 (HIPONOT01), 824975R1 (PROSNOT06), 903723R2 (COLNNOT07), 932897T1 (CERVNOT01), 1357827H1 (LUNGNOT09), 1504392F1 (BRAITUT07), 2618018F6 (GBLANOT01), 2668440H1 (ESOGTUT02), 2705719H1 (PONSAZT01), 2786769H1 (BRSTNOT13), 3256118H1 (OVARTUN01), 5734061H1 (KIDCTMT01)
10	20	4759250	BRAMNOT01	478443R1 (MMLR2DT01), 1290314T1 (BRAINOT11), 1713376F6 (UCMCNOT02), 2060032R6 (OVARNOT03), 3074325H1 (BONEUNT01), 3579609H1 (293TF3T01), 4289223H1 (BRABDIR01)

Table 2

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
1	243	S39 S73 S214 S53	N229	ATP/GTP-binding site motif A (P-loop): G49-T56 Phosphoenolpyruvate carboxykinase (ATP) (EC 4.1.1.49) signature: Y45-A106		MOTIFS ProfileScan
2	425	S47 S58 T121 S206 T335 S349 T79 S89 T110 T128 T180 Y248	N39 N45 N321	NAD dependent epimerase/ dehydratase family domains: L97-S107, D159-K196, R201-G212, S297-D309, A359-E399 Signal peptide: M1-G35 Transmembrane domain: M18-F37	Similar to thymidine diphosphoglucose 4,6-dehydratase (EC 4.2.1.46) [Caenorhabditis elegans] g1065948	BLAST-GenBank BLIMPS-BLOCKS HMMER MOTIFS SPScan
3	216	T74 S6 S9 T12 S182	N180	Isocitrate lyase (EC 4.1.3.1) signature: K16-T61		MOTIFS ProfileScan

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
4	343	S63 S126 S200 T319 T94 S257	N61 N191	Class II aldolase (EC 4.1.2.13) signature: F215-S236 Rlu family of pseudouridine synthases (EC 4.2.1.70) signature: V75-G95, V123-A165, T233-S257	Pseudouridylate synthase (EC 4.2.1.70) [Escherichia coli] g1786244 (P=3.9e-6)	BLIMPS-BLOCKS BLIMPS-PFAM MOTIFS
5	74	T12 T30 T4		Fumarate lyase signature: E33-I74		MOTIFS ProfileScan
6	176	T94 S144 T25		Class II aldolase (EC 4.1.2.13) signature: V77-T94 Glyoxalase I (EC 4.4.1.5) signature: H50-F88, M96-D105, G119-A133 Hypothetical YQJC protein; lyase domain: L45-L173 Signal peptide: M1-A28	Similarity to YQJC protein [C. elegans] g3875398	BLAST-GenBank BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PFAM MOTIFS SPScan

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
7	374	S60 T126 S165 S203 T283 T303 S305 S307 S36 S37 S81 S89 T113 S182 S192 S330 Y140 Y326		NAD dependent epimerase/ dehydratase family domains: L45-S55, K91-H105, D117-H154, K160-G171, E198-I235, P223-N358, S249-D261, E277-E286, D318-N358 Signal peptide: M1-G28	Similar to dTDP-D-glucose 4,6-dehydratase (EC 4.2.1.46) [Arabidopsis thaliana] g4836876	BLAST-GenBank BLAST-PRODOM BLIMPS-BLOCKS HMMER MOTIFS SPScan
8	780	S35 T64 S388 S389 T491 T515 T708 T757 T761 S66 S243 T256 T310 T477 T504 S562 S631 T646 S669 Y432 Y513	N341 N387 N475 N612 N746 N755 N759	Aconitase family (EC 4.2.1.3) signature: L63-V490, T64-G503, G72-G269, L140-F153, V163-G171, K167-N180, Y183-P196, S193-V247, N197-D212, G259-V272, P270-L514, D273-M286, P347-V361, L357-L408, G380-S389, L381-D392, D430-G453, G440-G453, I468-A482, V588-Q780 Aconitase (EC 4.2.1.3) C-terminal domain: L582-I752	Aconitate hydratase (EC 4.2.1.3) [Homo sapiens] g3366620	BLAST-GenBank BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS ProfileScan

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
9	594	T55 S181 T240 S317 S344 T391 S403 T428 T443 T444 S513 T563 S565 S584 T9 S21 S133 T213 S346 T440 S458 S574 Y310	N27 N278 N331 N561 N567		Ribulose-1,5-bisphosphate carboxylase (EC 4.1.1.39)/oxygenase small subunit N-methyltransferase I (EC 2.1.1.43) [Spinacia oleracea] g3403236	BLAST-GenBank MOTIFS
10	298	S265 T19	N129	Glyoxalase I (EC 4.4.1.5) signature: H8-N46, A69-G78, V242-C254 Glyoxalase (EC 4.4.1.5) domain: K12-R130	Similar to glyoxalase (EC 4.4.1.5) [Caenorhabditis elegans] g3874388	BLAST-GenBank BLIMPS-BLOCKS HMMER-PFAM MOTIFS

Table 3

Nucleotide SEQ ID NO:	Selected Fragment(s)	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
11	381-425	Reproductive (0.323) Nervous (0.194)	Cancer (0.452) Inflammation/Trauma (0.291) Cell Proliferation (0.129)	PBLUESCRIPT
12	140-187	Reproductive (0.255) Cardiovascular (0.236) Nervous (0.164)	Cancer (0.600) Inflammation/Trauma (0.237) Cell Proliferation (0.200)	pINCY
13	433-477 757-801	Reproductive (0.202) Cardiovascular (0.167) Hematopoietic/Immune (0.155) Nervous (0.155)	Inflammation/Trauma (0.405) Cancer (0.381) Cell Proliferation (0.179)	PSPORT1
14	434-478 893-937	Reproductive (0.188) Hematopoietic/Immune (0.167) Cardiovascular (0.146)	Cancer (0.479) Inflammation/Trauma (0.312) Cell Proliferation (0.188)	pINCY
15	416-460	Hematopoietic/Immune (0.167) Nervous (0.167) Urologic (0.167)	Cancer (0.444) Cell Proliferation (0.389) Inflammation/Trauma (0.167)	pINCY
16	1-45 595-639	Reproductive (0.400) Cardiovascular (0.160) Hematopoietic/Immune (0.120) Gastrointestinal (0.120)	Cancer (0.480) Inflammation/Trauma (0.280) Cell Proliferation (0.120)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragment(s)	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
17	11-55	Reproductive (0.222) Hematopoietic/Immune (0.194) Cardiovascular (0.167)	Cancer (0.500) Inflammation/Trauma (0.305) Cell Proliferation (0.139)	pINCY
18	1-45	Reproductive (0.233) Nervous (0.210) Gastrointestinal (0.152)	Cancer (0.443) Inflammation/Trauma (0.343) Cell Proliferation (0.171)	pINCY
19	164-208 758-802 1028-1072	Reproductive (0.269) Nervous (0.202) Gastrointestinal (0.151)	Cancer (0.429) Inflammation/Trauma (0.353) Cell Proliferation (0.176)	PSPORT1
20	1-46	Reproductive (0.256) Nervous (0.186) Gastrointestinal (0.163)	Cancer (0.434) Inflammation/Trauma (0.357) Cell Proliferation (0.209)	pINCY

Table 4

Nucleotide SEQ ID NO:	Library	Library Comment
11	LIVRNOT01	Library was constructed at Stratagene, using RNA isolated from the liver tissue of a 49-year-old male.
12	LUNGFET03	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
13	BRAITUT02	Library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney.
14	TESTTUT02	Library was constructed using RNA isolated from testicular tumor removed from a 31-year-old Caucasian male during unilateral orchiectomy. Pathology indicated embryonal carcinoma.
15	THYRNOT09	Library was constructed using RNA isolated from diseased thyroid tissue removed from an 18-year-old Caucasian female during an unilateral thyroid lobectomy and regional lymph node excision. Pathology indicated adenomatous goiter. This was associated with a follicular adenoma of the thyroid. Family history included thyroid cancer in the father.
16	SINIUCT01	Library was constructed using RNA isolated from ileum tissue obtained from a 42-year-old Caucasian male during a total intra-abdominal colectomy and endoscopic jejunostomy. Previous surgeries included polypectomy, colonoscopy, and spinal canal exploration. Family history included cerebrovascular disease, benign hypertension, atherosclerotic coronary artery disease, and type II diabetes.
17	PENCNOT01	Library was constructed using RNA isolated from penis corpus cavernosum tissue removed from a 53-year-old male. Patient history included untreated penile carcinoma.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Comment
18	LUNGTUT12	Library was constructed using RNA isolated from tumorous lung tissue removed from a 70-year-old Caucasian female during a lung lobectomy of the left upper lobe. Pathology indicated grade 3 (of 4) adenocarcinoma and vascular invasion. Patient history included tobacco abuse, depressive disorder, anxiety state, and skin cancer. Family history included cerebrovascular disease, congestive heart failure, colon cancer, depressive disorder, and primary liver cancer.
19	OVARTUN01	Library was constructed from RNA isolated from tumor tissue removed from the left ovary of a 58-year-old Caucasian female during a total abdominal hysterectomy, removal of a single ovary, and inguinal hernia repair. Pathology indicated a metastatic grade 3 adenocarcinoma of colonic origin, forming a partially cystic and necrotic tumor mass in the left ovary, and a nodule in the left mesovarium. A single intramural leiomyoma was identified in the myometrium. The cervix showed mild chronic cystic cervicitis. Patient history included benign hypertension, follicular ovarian cyst, colon cancer, benign colon neoplasm, and osteoarthritis. Family history included emphysema, myocardial infarction, atherosclerotic coronary artery disease, benign hypertension, hyperlipidemia, and primary tuberculous complex. The library was normalized and hybridized under conditions adapted from Soares et al. (Proc. Natl. Acad. Sci. USA (1994) 91:9228) and Bonaldo et al. (Genome Research (1996) 6:791).
20	BRAMNOT01	Library was constructed using RNA isolated from medulla tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. In addition, scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-10.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:11-20.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,

b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

18. A method for treating a disease or condition associated with decreased expression of functional HLYAP, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional HLYAP, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional HLYAP, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

<110> INCYTE GENOMICS, INC.
YUE, Henry
BANDMAN, Olga
TANG, Y. Tom
HILLMAN, Jennifer L.
LU, Dyung Aina M.
BAUGHN, Mariah R.

<120> HUMAN LYASES AND ASSOCIATED PROTEINS

<130> PF-0759 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/172,307
<151> 1999-12-16

<160> 20

<170> PERL Program

<210> 1
<211> 243
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 168714

<400> 1
Met Glu Asp Ser Phe Leu Gln Ser Phe Gly Arg Leu Ser Leu Gln
1 5 10 15
Pro Gln Gln Gln Gln Gln Arg Gln Arg Pro Pro Arg Pro Pro Pro
20 25 30
Arg Gly Thr Pro Pro Arg Arg His Ser Phe Arg Lys His Leu Tyr
35 40 45
Leu Leu Arg Gly Leu Pro Gly Ser Gly Lys Thr Thr Leu Ala Arg
50 55 60
Gln Leu Gln His Asp Phe Pro Arg Ala Leu Ile Phe Ser Thr Asp
65 70 75
Asp Phe Phe Phe Arg Glu Asp Gly Ala Tyr Glu Phe Asn Pro Asp
80 85 90
Phe Leu Glu Glu Ala His Glu Trp Asn Gln Lys Arg Ala Arg Lys
95 100 105
Ala Met Arg Asn Gly Ile Ser Pro Ile Ile Ile Asp Asn Thr Asn
110 115 120
Leu His Ala Trp Glu Met Lys Pro Tyr Ala Val Met Ala Leu Glu
125 130 135
Asn Asn Tyr Glu Val Ile Phe Arg Glu Pro Asp Thr Arg Trp Lys
140 145 150
Phe Asn Val Gln Glu Leu Ala Arg Arg Asn Ile His Gly Val Ser
155 160 165
Arg Glu Lys Ile His Arg Met Lys Glu Arg Tyr Glu His Asp Val
170 175 180
Thr Phe His Ser Val Leu His Ala Glu Lys Pro Ser Arg Met Asn
185 190 195
Arg Asn Gln Asp Arg Asn Asn Ala Leu Pro Ser Asn Asn Ala Arg
200 205 210
Tyr Trp Asn Ser Tyr Thr Glu Phe Pro Asn Arg Arg Ala His Gly
215 220 225
Gly Phe Thr Asn Glu Ser Ser Tyr His Arg Arg Gly Gly Cys His
230 235 240
His Gly Tyr

<210> 2
<211> 425
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1851727

<400> 2
Met Val Ser Lys Ala Leu Leu Arg Leu Val Ser Ala Val Asn Arg
1 5 10 15
Arg Arg Met Lys Leu Leu Leu Gly Ile Ala Leu Leu Ala Tyr Val
20 25 30
Ala Ser Val Trp Gly Asn Phe Val Asn Met Ser Phe Leu Leu Asn
35 40 45
Arg Ser Ile Gln Glu Asn Gly Glu Leu Lys Ile Glu Ser Lys Ile
50 55 60
Glu Glu Met Val Glu Pro Leu Arg Glu Lys Ile Arg Asp Leu Glu
65 70 75
Lys Ser Phe Thr Gln Lys Tyr Pro Pro Val Lys Phe Leu Ser Glu
80 85 90
Lys Asp Arg Lys Arg Ile Leu Ile Thr Gly Gly Ala Gly Phe Val
95 100 105
Gly Ser His Leu Thr Asp Lys Leu Met Met Asp Gly His Glu Val
110 115 120
Thr Val Val Asp Asn Phe Phe Thr Gly Arg Lys Arg Asn Val Glu
125 130 135
His Trp Ile Gly His Glu Asn Phe Glu Leu Ile Asn His Asp Val
140 145 150
Val Glu Pro Leu Tyr Ile Glu Val Asp Gln Ile Tyr His Leu Ala
155 160 165
Ser Pro Ala Ser Pro Pro Asn Tyr Met Tyr Asn Pro Ile Lys Thr
170 175 180
Leu Lys Thr Asn Thr Ile Gly Thr Leu Asn Met Leu Gly Leu Ala
185 190 195
Lys Arg Val Gly Ala Arg Leu Leu Leu Ala Ser Thr Ser Glu Val
200 205 210
Tyr Gly Asp Pro Glu Val His Pro Gln Ser Glu Asp Tyr Trp Gly
215 220 225
His Val Asn Pro Ile Gly Pro Arg Ala Cys Tyr Asp Glu Gly Lys
230 235 240
Arg Val Ala Glu Thr Met Cys Tyr Ala Tyr Met Lys Gln Glu Gly
245 250 255
Val Glu Val Arg Val Ala Arg Ile Phe Asn Thr Phe Gly Pro Arg
260 265 270
Met His Met Asn Asp Gly Arg Val Val Ser Asn Phe Ile Leu Gln
275 280 285
Ala Leu Gln Gly Glu Pro Leu Thr Val Tyr Gly Ser Gly Ser Gln
290 295 300
Thr Arg Ala Phe Gln Tyr Val Ser Asp Leu Val Asn Gly Leu Val
305 310 315
Ala Leu Met Asn Ser Asn Val Ser Ser Pro Val Asn Leu Gly Asn
320 325 330
Pro Glu Glu His Thr Ile Leu Glu Phe Ala Gln Leu Ile Lys Asn
335 340 345
Leu Val Gly Ser Gly Ser Glu Ile Gln Phe Leu Ser Glu Ala Gln
350 355 360
Asp Asp Pro Gln Lys Arg Lys Pro Asp Ile Lys Lys Ala Lys Leu
365 370 375
Met Leu Gly Trp Glu Pro Val Val Pro Leu Glu Glu Gly Leu Asn
380 385 390
Lys Ala Ile His Tyr Phe Arg Lys Glu Leu Glu Tyr Gln Ala Asn
395 400 405
Asn Gln Tyr Ile Pro Lys Pro Lys Pro Ala Arg Ile Lys Lys Gly
410 415 420
Arg Thr Arg His Ser

425

<210> 3
<211> 216
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2095185

<400> 3
Met Ser Phe Leu Phe Ser Ser Arg Ser Ser Lys Thr Phe Lys Pro
1 5 10 15
Lys Lys Asn Ile Pro Glu Gly Ser His Gln Tyr Glu Leu Leu Lys
20 25 30
His Ala Glu Ala Thr Leu Gly Ser Gly Asn Leu Arg Gln Ala Val
35 40 45
Met Leu Pro Glu Gly Glu Asp Leu Asn Glu Trp Ile Ala Val Asn
50 55 60
Thr Val Asp Phe Phe Asn Gln Ile Asn Met Leu Tyr Gly Thr Ile
65 70 75
Thr Glu Phe Cys Thr Glu Ala Ser Cys Pro Val Met Ser Ala Gly
80 85 90
Pro Arg Tyr Glu Tyr His Trp Ala Asp Gly Thr Asn Ile Lys Lys
95 100 105
Pro Ile Lys Cys Ser Ala Pro Lys Tyr Ile Asp Tyr Leu Met Thr
110 115 120
Trp Val Gln Asp Gln Leu Asp Asp Glu Thr Leu Phe Pro Ser Lys
125 130 135
Ile Gly Val Pro Phe Pro Lys Asn Phe Met Ser Val Ala Lys Thr
140 145 150
Ile Leu Lys Arg Leu Phe Arg Val Tyr Ala His Ile Tyr His Gln
155 160 165
His Phe Asp Ser Val Met Gln Leu Gln Glu Glu Ala His Leu Asn
170 175 180
Thr Ser Phe Lys His Phe Ile Phe Phe Val Gln Glu Phe Asn Leu
185 190 195
Ile Asp Arg Arg Glu Leu Ala Pro Leu Gln Glu Leu Ile Glu Lys
200 205 210
Leu Gly Ser Lys Asp Arg
215

<210> 4
<211> 343
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2342959

<400> 4
Met Asp Gly Arg Arg Val Leu Gly Arg Phe Trp Ser Gly Trp Arg
1 5 10 15
Arg Gly Leu Gly Val Arg Pro Val Pro Glu Asp Ala Gly Phe Gly
20 25 30
Thr Glu Ala Arg His Gln Arg Gln Pro Arg Gly Ser Cys Gln Arg
35 40 45
Ser Gly Pro Leu Gly Asp Gln Pro Phe Ala Gly Leu Leu Pro Lys
50 55 60
Asn Leu Ser Arg Glu Glu Leu Val Asp Ala Leu Arg Ala Ala Val
65 70 75
Val Asp Arg Lys Gly Pro Leu Val Thr Leu Asn Lys Pro Gln Gly
80 85 90

```
Leu Pro Val Thr Gly Lys Pro Gly Glu Leu Thr Leu Phe Ser Val
                95                  100                 105
Leu Pro Glu Leu Ser Gln Ser Leu Gly Leu Arg Glu Gln Glu Leu
            110                 115                 120
Gln Val Val Arg Ala Ser Gly Lys Glu Ser Ser Gly Leu Val Leu
        125                 130                 135
Leu Ser Ser Cys Pro Gln Thr Ala Ser Arg Leu Gln Lys Tyr Phe
    140                 145                 150
Thr His Ala Arg Arg Ala Gln Arg Pro Thr Ala Thr Tyr Cys Ala
155                 160                 165
Val Thr Asp Gly Ile Pro Ala Ala Ser Glu Gly Lys Ile Gln Ala
            170                 175                 180
Ala Leu Lys Leu Glu His Ile Asp Gly Val Asn Leu Thr Val Pro
        185                 190                 195
Val Lys Ala Pro Ser Arg Lys Asp Ile Leu Glu Gly Val Lys Lys
    200                 205                 210
Thr Leu Ser His Phe Arg Val Val Ala Thr Gly Ser Gly Cys Ala
215                 220                 225
Leu Val Gln Leu Gln Pro Leu Thr Val Phe Ser Ser Gln Leu Gln
            230                 235                 240
Val His Met Val Leu Gln Leu Cys Pro Val Leu Gly Asp His Met
        245                 250                 255
Tyr Ser Ala Arg Val Gly Thr Val Leu Gly Gln Arg Phe Leu Leu
    260                 265                 270
Pro Ala Glu Asn Asn Lys Pro Gln Arg Gln Val Leu Asp Glu Ala
275                 280                 285
Leu Leu Arg Arg Leu His Leu Thr Pro Ser Gln Ala Ala Gln Leu
            290                 295                 300
Pro Leu His Leu His Leu His Arg Leu Leu Leu Pro Gly Thr Arg
        305                 310                 315
Ala Arg Asp Thr Pro Val Glu Leu Leu Ala Pro Leu Pro Pro Tyr
            320                 325                 330
Phe Ser Arg Thr Leu Gln Cys Leu Gly Leu Arg Leu Gln
        335                 340
```

<210> 5
<211> 74
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2613975

<400> 5

```
Met Ser Asp Thr Arg Arg Arg Val Lys Val Tyr Thr Leu Asn Glu
  1               5                  10                  15
Asp Arg Gln Trp Asp Asp Arg Gly Thr Gly His Val Ser Ser Thr
            20                  25                  30
Tyr Val Glu Glu Leu Lys Gly Met Ser Leu Leu Val Arg Ala Glu
        35                  40                  45
Ser Asp Gly Ser Leu Leu Leu Glu Ser Lys Ile Asn Pro Asn Thr
    50                  55                  60
Ala Tyr Gln Lys Gln Gln Ala Ser Ser Cys Leu Ser Leu Ile
65                  70
```

<210> 6
<211> 176
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2683534

<400> 6
Met Ala Arg Val Leu Lys Ala Ala Ala Ala Asn Ala Val Gly Leu
1 5 10 15
Phe Ser Arg Leu Gln Ala Pro Ile Pro Thr Val Arg Ala Ser Ser
20 25 30
Thr Ser Gln Pro Leu Asp Gln Val Thr Gly Ser Val Trp Asn Leu
35 40 45
Gly Arg Leu Asn His Val Ala Ile Ala Val Pro Asp Leu Glu Lys
50 55 60
Ala Ala Ala Phe Tyr Lys Asn Ile Leu Gly Ala Gln Val Ser Glu
65 70 75
Ala Val Pro Leu Pro Glu His Gly Val Ser Val Val Phe Val Asn
80 85 90
Leu Gly Asn Thr Lys Met Glu Leu Leu His Pro Leu Gly Arg Asp
95 100 105
Ser Pro Ile Ala Gly Phe Leu Gln Lys Asn Lys Ala Gly Gly Met
110 115 120
His His Ile Cys Ile Glu Val Asp Asn Ile Asn Ala Ala Val Met
125 130 135
Asp Leu Lys Lys Lys Lys Ile Arg Ser Leu Ser Glu Glu Val Lys
140 145 150
Ile Gly Ala His Gly Lys Pro Val Ile Phe Leu His Pro Lys Asp
155 160 165
Cys Gly Gly Val Leu Val Glu Leu Glu Gln Ala
170 175

<210> 7
<211> 374
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2801723

<400> 7
Met Val Ala Ala Glu Leu Pro Cys Ala Phe Gln Thr Ile Leu Phe
1 5 10 15
Thr Val Leu Gly Thr Ala Glu Leu Gly Asp Val Gly Gly Val Leu
20 25 30
Gly Gly Thr Val Gly Ser Ser Arg Arg Leu Cys Glu Arg Val Leu
35 40 45
Val Thr Gly Gly Ala Gly Phe Ile Ala Ser His Met Ile Val Ser
50 55 60
Leu Val Glu Asp Tyr Pro Asn Tyr Met Ile Ile Asn Leu Asp Lys
65 70 75
Leu Asp Tyr Cys Ala Ser Leu Lys Asn Leu Glu Thr Ile Ser Asn
80 85 90
Lys Gln Asn Tyr Lys Phe Ile Gln Gly Asp Ile Cys Asp Ser His
95 100 105
Phe Val Lys Leu Leu Phe Glu Thr Glu Lys Ile Asp Ile Val Leu
110 115 120
His Phe Ala Ala Gln Thr His Val Asp Leu Ser Phe Val Arg Ala
125 130 135
Phe Glu Phe Thr Tyr Val Asn Val Tyr Gly Thr His Val Leu Val
140 145 150
Ser Ala Ala His Glu Ala Arg Val Glu Lys Phe Ile Tyr Val Ser
155 160 165
Thr Asp Glu Val Tyr Gly Gly Ser Leu Asp Lys Glu Phe Asp Glu
170 175 180
Ser Ser Pro Lys Gln Pro Thr Asn Pro Tyr Ala Ser Ser Lys Ala
185 190 195
Ala Ala Glu Cys Phe Val Gln Ser Tyr Trp Glu Gln Tyr Lys Phe
200 205 210
Pro Val Val Ile Thr Arg Ser Ser Asn Val Tyr Gly Pro His Gln
215 220 225

```
Tyr Pro Glu Lys Val Ile Pro Lys Phe Ile Ser Leu Leu Gln His
                230                 235                 240
Asn Arg Lys Cys Cys Ile His Gly Ser Gly Leu Gln Thr Arg Asn
                245                 250                 255
Phe Leu Tyr Ala Thr Asp Val Val Glu Ala Phe Leu Thr Val Leu
                260                 265                 270
Lys Lys Gly Lys Pro Gly Glu Ile Tyr Asn Ile Gly Thr Asn Phe
                275                 280                 285
Glu Met Ser Val Val Gln Leu Ala Lys Glu Leu Ile Gln Leu Ile
                290                 295                 300
Lys Glu Thr Asn Ser Glu Ser Glu Met Glu Asn Trp Val Asp Tyr
                305                 310                 315
Val Asn Asp Arg Pro Thr Asn Asp Met Arg Tyr Pro Met Lys Ser
                320                 325                 330
Glu Lys Ile His Gly Leu Gly Trp Arg Pro Lys Val Pro Trp Lys
                335                 340                 345
Glu Gly Ile Lys Lys Thr Ile Glu Trp Tyr Arg Glu Asn Phe His
                350                 355                 360
Asn Trp Lys Asn Val Glu Lys Ala Leu Glu Pro Phe Pro Val
                365                 370
```

<210> 8
<211> 780
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3130234

<400> 8

```
Met Ala Pro Tyr Ser Leu Leu Val Thr Arg Leu Gln Lys Ala Leu
  1                 5                  10                  15
Gly Val Arg Gln Tyr His Val Ala Ser Val Leu Cys Gln Arg Ala
                 20                  25                  30
Lys Val Ala Met Ser His Phe Glu Pro Asn Glu Tyr Ile His Tyr
                 35                  40                  45
Asp Leu Leu Glu Lys Asn Ile Asn Ile Val Arg Lys Arg Leu Asn
                 50                  55                  60
Arg Pro Leu Thr Leu Ser Glu Lys Ile Val Tyr Gly His Leu Asp
                 65                  70                  75
Asp Pro Ala Ser Gln Glu Ile Glu Arg Gly Lys Ser Tyr Leu Arg
                 80                  85                  90
Leu Arg Pro Asp Arg Val Ala Met Gln Asp Ala Thr Ala Gln Met
                 95                 100                 105
Ala Met Leu Gln Phe Ile Ser Ser Gly Leu Ser Lys Val Ala Val
                110                 115                 120
Pro Ser Thr Ile His Cys Asp His Leu Ile Glu Ala Gln Val Gly
                125                 130                 135
Gly Glu Lys Asp Leu Arg Arg Ala Lys Asp Ile Asn Gln Glu Val
                140                 145                 150
Tyr Asn Phe Leu Ala Thr Ala Gly Ala Lys Tyr Gly Val Gly Phe
                155                 160                 165
Trp Lys Pro Gly Ser Gly Ile Ile His Gln Ile Ile Leu Glu Asn
                170                 175                 180
Tyr Ala Tyr Pro Gly Val Leu Leu Ile Gly Thr Asp Ser His Thr
                185                 190                 195
Pro Asn Gly Gly Gly Leu Gly Gly Ile Cys Ile Gly Val Gly Gly
                200                 205                 210
Ala Asp Ala Val Asp Val Met Ala Gly Ile Pro Trp Glu Leu Lys
                215                 220                 225
Cys Pro Lys Val Ile Gly Val Lys Leu Thr Gly Ser Leu Ser Gly
                230                 235                 240
Trp Ser Ser Pro Lys Asp Val Ile Leu Lys Val Ala Gly Ile Leu
                245                 250                 255
Thr Val Lys Gly Gly Thr Gly Ala Ile Val Glu Tyr His Gly Pro
```

```
                260                 265                 270
Gly Val Asp Ser Ile Ser Cys Thr Gly Met Ala Thr Ile Cys Asn
            275                 280                 285
Met Gly Ala Glu Ile Gly Ala Thr Thr Ser Val Phe Pro Tyr Asn
        290                 295                 300
His Arg Met Lys Lys Tyr Leu Ser Lys Thr Gly Arg Glu Asp Ile
    305                 310                 315
Ala Asn Leu Ala Asp Glu Phe Lys Asp His Leu Val Pro Asp Pro
320                 325                 330
Gly Cys His Tyr Asp Gln Leu Ile Glu Ile Asn Leu Ser Glu Leu
            335                 340                 345
Lys Pro His Ile Asn Gly Pro Phe Thr Pro Asp Leu Ala His Pro
        350                 355                 360
Val Ala Glu Val Gly Lys Val Ala Glu Lys Glu Gly Trp Pro Leu
    365                 370                 375
Asp Ile Arg Val Gly Leu Ile Gly Ser Cys Thr Asn Ser Ser Tyr
380                 385                 390
Glu Asp Met Gly Arg Ser Ala Ala Val Ala Lys Gln Ala Leu Ala
            395                 400                 405
His Gly Leu Lys Cys Lys Ser Gln Phe Thr Ile Thr Pro Gly Ser
        410                 415                 420
Glu Gln Ile Arg Ala Thr Ile Glu Arg Asp Gly Tyr Ala Gln Ile
    425                 430                 435
Leu Arg Asp Leu Gly Gly Ile Val Leu Ala Asn Ala Cys Gly Pro
440                 445                 450
Cys Ile Gly Gln Trp Asp Arg Lys Asp Ile Lys Lys Gly Glu Lys
            455                 460                 465
Asn Thr Ile Val Thr Ser Tyr Asn Arg Asn Phe Thr Gly Arg Asn
        470                 475                 480
Asp Ala Asn Pro Glu Thr His Ala Phe Val Thr Ser Pro Glu Ile
    485                 490                 495
Val Thr Ala Leu Ala Ile Ala Gly Thr Leu Lys Phe Asn Pro Glu
500                 505                 510
Thr Asp Tyr Leu Thr Gly Thr Asp Gly Lys Lys Phe Arg Leu Glu
            515                 520                 525
Ala Pro Asp Ala Asp Glu Leu Pro Lys Gly Glu Phe Asp Pro Gly
        530                 535                 540
Gln Asp Thr Tyr Gln His Pro Pro Lys Asp Ser Ser Gly Gln His
    545                 550                 555
Val Asp Val Ser Pro Thr Ser Gln Arg Leu Gln Leu Leu Glu Pro
560                 565                 570
Phe Asp Lys Trp Asp Gly Lys Asp Leu Glu Asp Leu Gln Ile Leu
            575                 580                 585
Ile Lys Val Lys Gly Lys Cys Thr Thr Asp His Ile Ser Ala Ala
        590                 595                 600
Gly Pro Trp Leu Lys Phe Arg Gly His Leu Asp Asn Ile Ser Asn
    605                 610                 615
Asn Leu Leu Ile Gly Ala Ile Asn Ile Glu Asn Gly Lys Ala Asn
620                 625                 630
Ser Val Arg Asn Ala Val Thr Gln Glu Phe Gly Pro Val Pro Asp
            635                 640                 645
Thr Ala Arg Tyr Tyr Lys Lys His Gly Ile Arg Trp Val Val Ile
        650                 655                 660
Gly Asp Glu Asn Tyr Gly Glu Gly Ser Ser Arg Glu His Ala Ala
    665                 670                 675
Leu Glu Pro Arg His Leu Gly Gly Arg Ala Ile Ile Thr Lys Ser
680                 685                 690
Phe Ala Arg Ile His Glu Thr Asn Leu Lys Lys Gln Gly Leu Leu
            695                 700                 705
Pro Leu Thr Phe Ala Asp Pro Ala Asp Tyr Asn Lys Ile His Pro
        710                 715                 720
Val Asp Lys Leu Thr Ile Gln Gly Leu Lys Asp Phe Thr Pro Gly
    725                 730                 735
Lys Pro Leu Lys Cys Ile Ile Lys His Pro Asn Gly Thr Gln Glu
740                 745                 750
Thr Ile Leu Leu Asn His Thr Phe Asn Glu Thr Gln Ile Glu Trp
            755                 760                 765
```

Phe Arg Ala Gly Ser Ala Leu Asn Arg Met Lys Glu Leu Gln Gln
770 775 780

<210> 9
<211> 594
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3256118

<400> 9
Met Gly Lys Lys Ser Arg Val Lys Thr Gln Lys Ser Gly Thr Gly
1 5 10 15
Ala Thr Ala Thr Val Ser Pro Lys Glu Ile Leu Asn Leu Thr Ser
20 25 30
Glu Leu Leu Gln Lys Cys Ser Ser Pro Ala Pro Gly Pro Gly Lys
35 40 45
Glu Trp Glu Glu Tyr Val Gln Ile Arg Thr Leu Val Glu Lys Ile
50 55 60
Arg Lys Lys Gln Lys Gly Leu Ser Val Thr Phe Asp Gly Lys Arg
65 70 75
Glu Asp Tyr Phe Pro Asp Leu Met Lys Trp Ala Ser Glu Asn Gly
80 85 90
Ala Ser Val Glu Gly Phe Glu Met Val Asn Phe Lys Glu Glu Gly
95 100 105
Phe Gly Leu Arg Ala Thr Arg Asp Ile Lys Ala Glu Glu Leu Phe
110 115 120
Leu Trp Val Pro Arg Lys Leu Leu Met Thr Val Glu Ser Ala Lys
125 130 135
Asn Ser Val Leu Gly Pro Leu Tyr Ser Gln Asp Arg Ile Leu Gln
140 145 150
Ala Met Gly Asn Ile Ala Leu Ala Phe His Leu Leu Cys Glu Arg
155 160 165
Ala Ser Pro Asn Ser Phe Trp Gln Pro Tyr Ile Gln Thr Leu Pro
170 175 180
Ser Glu Tyr Asp Thr Pro Leu Tyr Phe Glu Glu Asp Glu Val Arg
185 190 195
Tyr Leu Gln Ser Thr Gln Ala Ile His Asp Val Phe Ser Gln Tyr
200 205 210
Lys Asn Thr Ala Arg Gln Tyr Ala Tyr Phe Tyr Lys Val Ile Gln
215 220 225
Thr His Pro His Ala Asn Lys Leu Pro Leu Lys Asp Ser Phe Thr
230 235 240
Tyr Glu Asp Tyr Arg Trp Ala Val Ser Ser Val Met Thr Arg Gln
245 250 255
Asn Gln Ile Pro Thr Glu Asp Gly Ser Arg Val Thr Leu Ala Leu
260 265 270
Ile Pro Leu Trp Asp Met Cys Asn His Thr Asn Gly Leu Ile Thr
275 280 285
Thr Gly Tyr Asn Leu Glu Asp Asp Arg Cys Glu Cys Val Ala Leu
290 295 300
Gln Asp Phe Arg Ala Gly Glu Gln Ile Tyr Ile Phe Tyr Gly Thr
305 310 315
Arg Ser Asn Ala Glu Phe Val Ile His Ser Gly Phe Phe Phe Asp
320 325 330
Asn Asn Ser His Asp Arg Val Lys Ile Lys Leu Gly Val Ser Lys
335 340 345
Ser Asp Arg Leu Tyr Ala Met Lys Ala Glu Val Leu Ala Arg Ala
350 355 360
Gly Ile Pro Thr Ser Ser Val Phe Ala Leu His Phe Thr Glu Pro
365 370 375
Pro Ile Ser Ala Gln Leu Leu Ala Phe Leu Arg Val Phe Cys Met
380 385 390
Thr Glu Glu Glu Leu Lys Glu His Leu Leu Gly Asp Ser Ala Ile

```
                395                 400                 405
Asp Arg Ile Phe Thr Leu Gly Asn Ser Glu Phe Pro Val Ser Trp
                410                 415                 420
Asp Asn Glu Val Lys Leu Trp Thr Phe Leu Glu Asp Arg Ala Ser
                425                 430                 435
Leu Leu Leu Lys Thr Tyr Lys Thr Thr Ile Glu Glu Asp Lys Ser
                440                 445                 450
Val Leu Lys Asn His Asp Leu Ser Val Arg Ala Lys Met Ala Ile
                455                 460                 465
Lys Leu Arg Leu Gly Glu Lys Glu Ile Leu Glu Lys Ala Val Lys
                470                 475                 480
Ser Ala Ala Val Asn Arg Glu Tyr Tyr Arg Gln Gln Met Glu Glu
                485                 490                 495
Lys Ala Pro Leu Pro Lys Tyr Glu Glu Ser Asn Leu Gly Leu Leu
                500                 505                 510
Glu Ser Ser Val Gly Asp Ser Arg Leu Pro Leu Val Leu Arg Asn
                515                 520                 525
Leu Glu Glu Glu Ala Gly Val Gln Asp Ala Leu Asn Ile Arg Glu
                530                 535                 540
Ala Ile Ser Lys Ala Lys Ala Thr Glu Asn Gly Leu Val Asn Gly
                545                 550                 555
Glu Asn Ser Ile Pro Asn Gly Thr Arg Ser Glu Asn Glu Ser Leu
                560                 565                 570
Asn Gln Glu Ser Lys Arg Ala Val Glu Asp Ala Lys Gly Ser Ser
                575                 580                 585
Ser Asp Ser Thr Ala Gly Val Lys Glu
                590
```

<210> 10
<211> 298
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4759250

<400> 10

```
Met Ala Ala Arg Arg Ala Leu His Phe Val Phe Lys Val Gly Asn
  1                 5                  10                  15
Arg Phe Gln Thr Ala Arg Phe Tyr Arg Asp Val Leu Gly Met Lys
                 20                  25                  30
Val Leu Arg His Glu Glu Phe Glu Glu Gly Cys Lys Ala Ala Cys
                 35                  40                  45
Asn Gly Pro Tyr Asp Gly Lys Trp Ser Lys Thr Met Val Gly Phe
                 50                  55                  60
Gly Pro Glu Asp Asp His Phe Val Ala Glu Leu Thr Tyr Asn Tyr
                 65                  70                  75
Gly Val Gly Asp Tyr Lys Leu Gly Asn Asp Phe Met Gly Ile Thr
                 80                  85                  90
Leu Ala Ser Ser Gln Ala Val Ser Asn Ala Arg Lys Leu Glu Trp
                 95                 100                 105
Pro Leu Thr Glu Val Ala Glu Gly Val Phe Glu Thr Glu Ala Pro
                110                 115                 120
Gly Gly Tyr Lys Phe Tyr Leu Gln Asn Arg Ser Leu Pro Gln Ser
                125                 130                 135
Asp Pro Val Leu Lys Val Thr Leu Ala Val Ser Asp Leu Gln Lys
                140                 145                 150
Ser Leu Asn Tyr Trp Cys Asn Leu Leu Gly Met Lys Ile Tyr Glu
                155                 160                 165
Lys Asp Glu Glu Lys Gln Arg Ala Leu Leu Gly Tyr Ala Asp Asn
                170                 175                 180
Gln Cys Lys Leu Glu Leu Gln Gly Val Lys Gly Gly Val Asp His
                185                 190                 195
Ala Ala Ala Phe Gly Arg Ile Ala Phe Ser Cys Pro Gln Lys Glu
                200                 205                 210
```

```
Leu Pro Asp Leu Glu Asp Leu Met Lys Arg Glu Asn Gln Lys Ile
                215                 220                 225
Leu Thr Pro Leu Val Ser Leu Asp Thr Pro Gly Lys Ala Thr Val
                230                 235                 240
Gln Val Val Ile Leu Ala Asp Pro Asp Gly His Glu Ile Cys Phe
                245                 250                 255
Val Gly Asp Glu Ala Phe Arg Glu Leu Ser Lys Met Asp Pro Glu
                260                 265                 270
Gly Ser Lys Leu Leu Asp Asp Ala Met Ala Ala Asp Lys Ser Asp
                275                 280                 285
Glu Trp Phe Ala Lys His Asn Lys Pro Lys Ala Ser Gly
                290                 295
```

<210> 11
<211> 1686
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 168714

<400> 11
```
gggctgtgag tctcccagcg tccccagctt tccaggtagg gacgcccccc cccacgcagc   60
acggttccgg cgggtggaaa ggaggggctg ggctcccagc gcccgccccc ctatccatca  120
catggccgga gagtcacaaa aacaacagct ttggccaaga ccgtgacttc agtaaaggga  180
acccggggct ctcgcagcca gccctcctgc ccatggagga cagtttcctt caatcttttg  240
ggaggctgag cctccagccc cagcagcagc agcagcggca gcggccgccc cggccgcccc  300
cgcgggggac acctcctcgc cgccacagct ttaggaaaca cctctacctc ctgcgaggcc  360
tcccgggctc cgggaaaact acactggcca gacaattgca gcatgacttt cccagggccc  420
tgattttcag cacggatgat tttttcttca gggaagatgg tgcctatgag ttcaatcctg  480
acttcctgga ggaagctcat gaatggaacc aaaaaagagc aagaaaagca atgaggaatg  540
gcatatcccc cattattatt gataatacca acctccacgc ctgggaaatg aagccctatg  600
cagtcatggc acttgaaaat aactatgaag ttatattccg agaacctgac actcgctgga  660
aattcaacgt tcaagagtta gcaagaagaa acattcatgg tgtctcaaga gaaaaaatcc  720
accgaatgaa agaacggtat gaacacgatg ttacttttca cagtgtgctt catgcagaaa  780
agccaagcag aatgaacaga aaccaggaca ggaatgaatgc attgccttcc aacaatgcca  840
gatactggaa ttcctacaca gagtttccaa accggagggc ccacggtgga tttacaaatg  900
agagctccta tcacagaagg ggcggttgtc accatggata ttagaggcct atcttacagc  960
caggcagaat tttcctaagt cagtttctac ttcagttttt gttatttttt gttgcatttt 1020
agtcagagct ccaattccag tgtaaatagc tgaactcaaa agtttctgag caaagtcatt 1080
atattcactt tcttcaccaa aatttgttaa agtgcttcta tatgcatggt ctgatgctgg 1140
gaattctgca gatttgagta aacagtctct ttctctaggg taagaatttg aaaccaaaac 1200
ttgagaacac acccaagaat atatttacat aggttcatag atgaaataaa gtgtttatat 1260
tatataataag cttcagtacc atttgctctg aagtgatcta tttatttttt caggaaattc 1320
atctccatcg gtaaagttgg gaaggtggag agaagtggtg gggggcagga aaagttttag 1380
tgccattgct actttgataa tctatgtatc caaaaatgtg agatgtgcga ctcttatgat 1440
actgattttc ctttaatgtt aatatgccag aaagcataca tctaagggaa cattgtcctt 1500
caaagtagac actttgggaa gttatttctt tattttaatg atgtatcatt gttaaaaatg 1560
ctgtcaaatc cttaatagct acaggagcta ctgagggaaa tcagtgtcat tatttaaagt 1620
cacgccttgt gtttttacta ctttattcag caggattaaa cctgaataac ttttggctgt 1680
tgtgct                                                           1686
```

<210> 12
<211> 2053
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1851727

<400> 12
```
caggcgggcc cccgcgcggc agggccctgg acccgcgcgg ctcccgggga tggtgagcaa   60
ggcgctgctg cgcctcgtgt ctgccgtcaa ccgcaggagg atgaagctgc tgctgggcat  120
```

cgccttgctg gcctacgtcg cctctgtttg gggcaacttc gttaatatga gctttctact 180
caacaggtct atccaggaaa atggtgaact aaaaattgaa agcaagattg aagagatggt 240
tgaaccacta agagagaaaa tcagagattt agaaaaaagc tttacccaga aatacccacc 300
agtaaagttt ttatcagaaa aggatcggaa aagaattttg ataacaggag gcgcagggtt 360
cgtgggctcc catctaactg acaaactcat gatggacggc cacgaggtga ccgtggtgga 420
caatttcttc acgggcagga agagaaacgt ggagcactgg atcggacatg agaacttcga 480
gttgattaac cacgacgtgg tggagcccct ctacatcgag gttgaccaga tataccatct 540
ggcatctcca gcctcccctc caaactacat gtataatcct atcaagacat taaagaccaa 600
tacgattggg acattaaaca tgttggggct ggcaaaacga gtcggtgccc gtctgctcct 660
ggcctccaca tcggaggtgt atggagatcc tgaagtccac cctcaaagtg aggattactg 720
gggccacgtg aatccaatag gacctcgggc ctgctacgat gaaggcaaac gtgttgcaga 780
gaccatgtgc tatgcctaca tgaagcagga aggcgtggaa gtgcgagtgg ccagaatctt 840
caacaccttt gggccacgca tgcacatgaa cgatgggcga gtagtcagca acttcatcct 900
gcaggcgctc caggggagc cactcacggt atacggatcc gggtctcaga caagggcgtt 960
ccagtacgtc agcgatctag tgaatggcct cgtggctctc atgaacagca acgtcagcag 1020
cccggtcaac ctggggaacc cagaagaaca cacaatccta gaatttgctc agttaattaa 1080
aaaccttgtt ggtagcggaa gtgaaattca gtttctctcc gaagcccagg atgacccaca 1140
gaaaagaaaa ccagacatca aaaaagcaaa gctgatgctg gggtgggagc ccgtggtccc 1200
gctggaggaa ggtttaaaca aagcaattca ctacttccgt aaagaactcg agtaccaggc 1260
aaataatcag tacatcccca aaccaaagcc tgccagaata aagaaaggac ggactcgcca 1320
cagctgaact cctcactttt aggacacaag actaccattg tacacttgat gggatgtatt 1380
tttggctttt ttttgttgtc gtttaaagaa agactttaac aggtgtcatg aagaacaaac 1440
tggaatttca ttctgaagct tgctttaatg aaatggatgt gcctaaaagc tcccctcaaa 1500
aaactgcaga ttttgccttg cacttttga atctctcttt ttatgtaaaa tagcgtagat 1560
gcatctctgc gtattttcaa gtttttttat cttgctgtga gagcatatgt tgtgactgtc 1620
gttgacagtt ttatttactg gtttctttgt gaagctgaaa aggaacatta agcgggacaa 1680
aaaatgccga ttttatttat aaaagtgggt acttaataaa tgagtcgtta tactatgcat 1740
aaagaaaaat cctagcagta ttgtcaggtg gtggtgcgcc ggcattgatt ttagggcaga 1800
taaaagaatt ctgtgtgaga gctttatgtt tctcttttaa ttcagagttt ttccaaggtc 1860
tacttttgag ttgcaaactt gactttgaaa tattcctgtt ggtcatgatc aaggatattt 1920
gaaatcacta ctgtgttttg ctgcgtatct ggggcggggg caggttgggg ggcacaaagt 1980
taacatattc ttggttaacc atggttaaat atgctatttt aataaaatat tgaaactcaa 2040
aaaaaaaaaa aaa 2053

<210> 13
<211> 2490
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2095185

<400> 13
gctcgaggcg aggtggggta ggcgggcaag gcgggcgccg aggtttgcaa aggctcgcag 60
cggccagaaa cccggctccg agcggcggcg gcccggcttc cgctgcccgt gagctaagga 120
cggtccgctc cctctagcca gctccgaatc ctgatccagg cgggggccag gggcccctcg 180
cctccccctct gaggaccgaa gatgagcttc ctcttcagca gccgctcttc taaaacattc 240
aaaccaaaga agaatatccc tgaaggatct catcagtatg aactcttaaa acatgcagaa 300
gcaactctag gaagtgggaa tctgagacaa gctgttatgt tgcctgaggg agaggatctc 360
aatgaatgga ttgctgtgaa cactgtggat ttctttaacc agatcaacat gttatatgga 420
actatttcag aattctgcac tgaagcaagc tgtccagtca tgtctgcagg tccgagatat 480
gaatatcact gggcagatgg tactaatatt aaaaagccaa tcaaatgttc tgcaccaaaa 540
tacattgact atttgatgac ttgggttcaa gatcagcttg atgatgaaac tctttttcct 600
tctaagattg gtgtcccatt tcccaaaaac tttatgtctg tggcaaagac tattctaaag 660
cgtctgttca gggtttatgc ccatatttat caccagcact ttgattctgt gatgcagctg 720
caagaggagg cccacctcaa cacctccttt aagcacttta ttttctttgt tcaggagttt 780
aatctgattg ataggcgtga gctggcacct cttcaagaat taatagagaa acttggatca 840
aaagacagat aaatgtttct tctagaacac agttaccccc ttgcttcatc tattgctaga 900
actatcttat tgctatttgt tatagactag tgatacaaac tttaagaaaa caggataaaa 960
agatacccat tgcctgtgtc tactgataaa attatcccaa aggtaggttg gtgtgatagt 1020
ttccgagtaa gaccttaagg acacagccaa atcttaagta ctgtgtgacc actcttgttg 1080
ttatcagata gtcatacttg gttgtaatat gtgatggtta acctgtagct tataaattta 1140
cttatttatc ttttactcat ttactcagtc atttctttac aagaaaatga ttgaatctgt 1200
tttaggtgac agcacaatgg acattaagaa tttccatcaa taatttatga ataagtttcc 1260
agaacaaatt tcctaataac acaatcagat tggttttatt cttttatttt acgaataaaa 1320

aatgtatttt tcagtatcct tgagatttag aacatctgtg tcacttcaga taacatttta 1380
gtttcaagtt tgtatggtag tgtttttata gataagatac gtctatttt tcaaaattca 1440
tgattgcagt ttaaatcatc atatgacgtg tgggtgggag caaccaaagt tatttttaca 1500
gggactttat tttttgatct ttatttgaga ttgttttcat atctatctaa attattagga 1560
gtgtgtgtat cagaagtaat tttttaatgt cttctaagga tggtcttcca ggcttttaaa 1620
ctgaaaagct taattcagat agtagctttt ggctgagaaa aggaatccaa aatattaata 1680
aatttagatc tcaaaaccac tatttttatt atttcattat ttttcagagg ccttaaaatt 1740
ctggataaga gaatggagga aaatactcag agtacttgat tattttattt ccttttatta 1800
aaaaattact tctatgtttt tattgtctct tgagccttag ttaagagtag tgtagaaatg 1860
catgaacttc atcctaataa ggataaaact taaggaaaac cacaataaac catgaaggtg 1920
tacacatctt ataacacaga taaagttttg gtgtgctacc tattcttgag agagtgagtg 1980
agtgtatgtg tttaaaggaa acaaaatggg agaaataagt tttaaaaaaa tcctcatttt 2040
gttaatattc aaaagatgga ctgagcttcc acttgggttt tatcttgttt taattgtttt 2100
tgtatcaaaa cttgaaattc ctctatttct attgggatat aaaagccttc cccttcagtg 2160
aagaaaacat ttatttttta tttgattcct aggatttagt aaactctagc tgtctattta 2220
aaatgtactg aggcacaaca agtattatac tggaagactt gccaaactgg caaagcttta 2280
agttcatcag cattctatgt ggttcagagc tgtgattttt gcaaagtatt ttaccaacct 2340
cctcgatggc tttgataaag gttagatttg atgttttttt ttagatttat ttttcttact 2400
ccactaaact ataaagaaaa taattactta gaaactccat tttaaataat catttcctag 2460
aaattcttaa atatatacag aattttaaag 2490

<210> 14
<211> 1230
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2342959

<400> 14
gcgcgctgtc ctggctcggg agatggacgg ccgccgtgtt ttgggccggt tctggagtgg 60
ctggcggcgg ggcctgggtg tccgcccagt gcccgaggac gcaggctttg gcaccgaagc 120
ccggcatcag aggcaacccc gcggctcctg ccaacggtcg gggcccctcg gggaccagcc 180
cttcgcgggg ctgctgccaa aaaacctcag tcgggaggag ctggttgatg cgctgcgggc 240
agccgtggtg gaccggaaag gacctctagt gacgttgaac aagccacagg gtctaccagt 300
gacaggaaaa ccaggagagc tgacgttgtt ctcagtgctg ccagagctga gccagtccct 360
agggctcagg gagcaggagc ttcaggttgt ccgagcatct gggaaagaaa gctctgggct 420
tgtactcctc tccagctgtc cccagacagc tagtcgcctc cagaagtact tcacccatgc 480
acggagagcc caaaggccca cagccaccta ctgtgctgtc actgatggga tcccagctgc 540
ttctgagggg aagatccagg ctgccctgaa actggaacac attgatgggg tcaatctcac 600
agttccagtg aaggccccat cccgaaagga catcctggaa ggtgtcaaga agactctcag 660
tcactttcgt gtggtagcca caggctctgg ctgtgccctg gtccagctgc agccactgac 720
agtgttctcc agtcaactac aggtgcacat ggtactacag ctctgccctg tgcttgggga 780
ccacatgtac tctgcccgtg tgggcactgt cctgggccag cgatttctgc tgccagctga 840
gaacaacaag ccccaaagac aggtcctgga tgaagccctc ctcagacgcc tccacctgac 900
cccctcccag gctgcccagc tgcccttgca cctccaccta catcggctcc ttctcccagg 960
caccagggcc agggacaccc ctgttgagct cctggcacca ctgccccctt atttctccag 1020
gaccctacag tgcctggggc tccgcttaca atagtcctcc ctctgttcct gacccctca 1080
cacacactgg aaagtgaggg tgggggctct gcagtcagac aaacctaaga tcacatcctg 1140
gacaggccac ttgcttgctg tgtggcattg ggcaagtaac tttacctctc tggacttgtg 1200
ataataaaag ttcctacctc aaaaaaaaaa 1230

<210> 15
<211> 955
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2613975

<400> 15
ctcctcgcga gatgccgagc attccggcct gggaagcgcg tgcagaagcg gaggtgctgc 60
tcatgggact tgtcggccgc cgtagcccct gctaggacag cccgtgcgag cctgctggag 120

gaggaagaga aaggcagaga gagtcgggtt acaagatggc ggatctgtag tagttaccgc 180
ggcggcggga gagcaagcga gccctggggg gcaaagagac gggagagtgg gtgtatgcgc 240
gggtgaagtg agaggtaacg gggcctccgg gcggagaggc ctcagtggct cttgtcaccc 300
cttctcgcgg ctgaaccttt ggagccatgg tgaattcggg cctctccgaa gccgccgccg 360
ccgccaccgc cactactgcc tttaccgtct cctaagagtg aggagcgcgg acgaggtaag 420
cgaggaggcg gcggctagag cggtggagac agcagccacc atgtcggata cgcggcggcg 480
agtgaaggtc tataccctga acgaagaccg gcaatgggac gaccgaggca ccgggcacgt 540
ctcctccact tacgtggagg agctcaaggg gatgtcgctg ctggttcggg cagagtccga 600
cggatcacta ctcttggaat caaagataaa tccaaatact gcatatcaga aacaacaggc 660
aagtagttgt ttatctttaa tttgaaagac ttcatctgtg atcaaggaag tattaatctg 720
acaaaggtgg gaaagctttc ctgacaagaa aaaaacatgt ttggtaaaca aagatcatgt 780
gtatttctct tgcaggttaa aagtttcaga ctgaaaaaaa gtttttgtac tggtgataat 840
tatcattttt ggattgagcc actgtcggtt tattctaaga tgtatttatt agtattattt 900
aactgtagtt agccaagctc ttctatacct tgacatgaaa ccttttattc tgagt 955

<210> 16
<211> 849
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2683534

<400> 16
cgtggctagt cttgacgtgg cgggttgctt tccaaaatgg cgcgggtgct gaaggctgca 60
gccgcgaatg ccgtagggct tttttccaga cttcaagctc ccattccaac agtaagagct 120
tcttccacat cacagccctt ggatcaagtg acaggttctg tgtggaacct gggtcgactc 180
aaccatgtag ccatagcagt gccagatttg gaaaaggctg cagcatttta taagaatatt 240
ctgggggccc aggtaagtga agcggtccct cttcctgaac atggagtatc tgttgttttt 300
gtcaacctgg gaaataccaa gatggaactg cttcatccat tgggacgtga cagtccaatt 360
gcaggttttc tgcagaaaaa caaggctgga ggaatgcatc acatctgcat cgaggtggat 420
aatattaatg cagctgtgat ggatttgaaa aaaaagaaga tccgcagtct aagtgaagag 480
gtcaaaatag gagcacatgg aaaaccagtg atttttctcc atcctaaaga ctgtggtgga 540
gtccttgtgg aactggagca agcttgattt atatttgcaa gcaactaaat taattgacct 600
gaaaaagcct atcaaatact atcaaaatgt actatgacat tgagtccttc actgcttcca 660
tcatgtaaaa gttcacagtt aaagactgaa ttacagaaag attaaaatat atacatatat 720
aaatacataa atatgtatat tatttagatt aacaaacata tttgttaatt tgaatttgaa 780
gaaaatcttg attactaatt acttagggaa cattattaaa atcatataga aataaattat 840
tcctcttct 849

<210> 17
<211> 1919
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2801723

<400> 17
gctgcctctg gctgctctgt taacgtgtcc cgcgagcgag gcgcgtccga aaatggtcgc 60
ggcggaactt ccctgcgctt ttcagaccat actctttacg gtactaggca ctgctgagct 120
gggagatgtc ggcggcgtgt tgggaggaac cgtggggtct tcccggcggc tttgcgaacg 180
ggtcctggtg accggcggtg ctggtttcat tgcatcacat atgattgtct ctttagtgga 240
agattatcca aactatatga tcataaatct agacaagctg gattactgtg caagcttgaa 300
gaatcttgaa accatttcta acaaacagaa ctacaaattt atacagggtg acatatgtga 360
ttctcacttt gtgaaactgc tttttgaaac agagaaaata gatatagtac tacattttgc 420
cgcacaaaca catgtagatc tttcattcgt acgtgccttt gagtttacct atgttaatgt 480
ttatggcact cacgttttgg taagtgctgc tcatgaagcc agagtggaga agtttattta 540
tgtcagcaca gatgaagtat atggtggcag tcttgataag gaatttgatg aatcttcacc 600
caaacaacct acaaatcctt atgcatcatc taaagcagct gctgaatgtt ttgtacagtc 660
ttactgggaa caatataagt ttccagttgt catcacaaga agcagtaatg tttatggacc 720
acatcaatat ccagaaaagg ttattccaaa atttatatct ttgctacagc acaacaggaa 780
atgttgcatt catgggtcag ggcttcaaac aagaaacttc ctttatgcta ctgatgttgt 840

agaagcattt ctcactgtcc tcaaaaaagg gaaaccaggt gaaatttata acatcggaac 900
caattttgaa atgtcagttg tccagcttgc caaagaacta atacaactga tcaaagagac 960
caattcagag tctgaaatgg aaaattgggt tgattatgtt aatgatagac ccaccaatga 1020
catgagatac ccaatgaagt cagaaaaaat acatggctta ggatggagac ctaaagtgcc 1080
ttggaaagaa ggaataaaga aaacaattga atggtacaga gagaattttc acaactggaa 1140
gaatgtggaa aaggcattag aacccttcc ggtataatca ccatttatat agtcgagaca 1200
gttgtcaaag aagaaagtta tcctacctcg ccaagtggta tgaaattaag tgaccaaatg 1260
aagtgcactc ttttcttttg gaattagatt catgactttc tgtataaaat tcaaatgcag 1320
aatgcctcaa tctttgggag agtttcagta ctggcataga atttaaatgt caaaattctt 1380
tctgaaaccc tttctcctag aaactaggaa ataataggtg tagaagactc tccctaaggg 1440
tagccaggaa gaagtctcct gattcggaca accatgaggg gtagtggtgc tagggagaag 1500
gcaaccttca ctggttttga actcagtgcc taagaaagtc tctgaaatgt tcgtttttag 1560
gcaatatagg atgtcttagg ccctaattca ccatttcttt tttaagatct gatatgctat 1620
cattgcctta ataatggaac aaaatagaag catatctaac actttttaaa ttgataattt 1680
tgtaaaattg attacgttga atgcttttta agagaagtgt gtaaagtttt tatattttca 1740
caattaacgt atgtaaaacc ttgtatcaga aatttatcat gtttactgtt taaaatgatt 1800
gtatttataa aattgtcaat atcttaatgt atttaatgta gaatattgct ttttaaaata 1860
atgtttttat tttgctgtag aaaaataaaa aaaaatttga ttataaaaaa aaaaaaaaa 1919

<210> 18
<211> 2735
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3130234

<400> 18
ctttgtcagt gcacaaaatg gcgccctaca gcctactggt gactcggctg cagaaagctc 60
tgggtgtgcg gcagtaccat gtggcctcag tcctgtgcca acgggccaag gtggcgatga 120
gccactttga gcccaacgag tacatccatt atgacctgct agagaagaac attaacattg 180
ttcgcaaacg actgaaccgg ccgctgacac tctcggagaa gattgtgtat ggacacctgg 240
atgaccccgc cagccaggaa attgagcgag gcaagtcgta cctgcggctg cggccggacc 300
gtgtggccat gcaggatgcg acggcccaga tggccatgct ccagttcatc agcagcgggc 360
tgtccaaggt ggctgtgcca tccaccatcc actgtgacca tctgattgaa gcccaggttg 420
ggggcgagaa agacctgcgc cgggccaagg acatcaacca ggaagtttat aatttcctgg 480
caactgcagg tgccaaatat ggcgtgggct tctggaagcc tggatctgga atcattcacc 540
agattattct ggaaaactat gcgtaccctg gtgttcttct gattggcact gactcccaca 600
cccccaatgg tggcggcctt gggggcatct gcattggagt tgggggtgcc gatgctgtgg 660
atgtcatggc tgggatcccc tgggagttga agtgccccaa ggtgattggc gtgaagctga 720
cgggctctct ctccggttgg tcctcaccca aagatgtgat cctgaaggtg gcaggcatcc 780
tcacggtgaa aggtggcaca ggtgcaatcg tggaatacca cgggcctggt gtagactcca 840
tctcctgcac tggcatggcg acaatctgca acatgggtgc agaaattggg gccaccactt 900
ccgtgttccc ttacaaccac aggatgaaga agtacctgag caagaccggc cgggaagaca 960
ttgccaatct agctgatgaa ttcaaggatc acttggtgcc tgaccctggc tgccattatg 1020
accaactaat tgaaattaac ctcagtgagc tgaagccaca catcaatggg cccttcaccc 1080
ctgacctggc tcaccctgtg gcagaagtgg gcaaggtggc agagaaggaa ggatggcctc 1140
tggacatccg agtgggtcta attggtagct gcaccaattc aagctatgaa gatatggggc 1200
gctcagcagc tgtggccaag caggcactgg cccatggcct caagtgcaag tcccagttca 1260
ccatcactcc aggttccgag cagatccgcg ccaccattga gcgggacggc tatgcacaga 1320
tcttgaggga tctgggtggc attgtcctgg ccaatgcttg tggcccctgc attggccagt 1380
gggacaggaa ggacatcaag aaggggggaga agaacacaat cgtcacctcc tacaacagga 1440
acttcacggg ccgcaacgac gcaaaccccg agacccatgc ctttgtcacg tccccagaga 1500
ttgtcacagc cctggccatt gcgggaaccc tcaagttcaa cccagagacc gactacctga 1560
cgggcacgga tggcaagaag ttcaggctgg aggctccgga tgcagatgag cttcccaaag 1620
gggagtttga cccagggcag gacacctacc agcacccacc caaggacagc agcgggcagc 1680
atgtggacgt gagccccacc agccagcgcc tgcagctcct ggagcctttt gacaagtggg 1740
atggcaagga cctggaggac ctgcagatcc tcatcaaggt caaagggaag tgtaccactg 1800
accacatctc agctgctggc ccctggctca agttccgtgg gcacttggat aacatctcca 1860
acaacctgct cattggtgcc atcaacattg aaaacggcaa ggccaactcc gtgcgcaatg 1920
ccgtcactca ggagtttggc cccgtccctg acactgcccg ctactacaag aaacatggca 1980
tcaggtgggt ggtgatcgga gacgagaact acggcgaggg ctcgagccgg gagcatgcag 2040
ctctggagcc tcgccacctt gggggccggg ccatcatcac caagagcttt gccaggatcc 2100
acgagaccaa cctgaagaaa cagggcctgc tgcctctgac cttcgctgac ccggctgact 2160
acaacaagat tcaccctgtg gacaagctga ccattcaggg cctgaaggac ttcacccctg 2220

gcaagcccct gaagtgcatc atcaagcacc ccaacgggac ccaggagacc atcctcctga 2280
accacacctt caacgagacg cagattgagt ggttccgcgc tggcagtgcc ctcaacagaa 2340
tgaaggaact gcaacagtga gggcagtgcc tccccgcccc gccgctggcg tcaagttcag 2400
ctccacgtgt gccatcagtg gatccgatcc gtccagccat ggcttcctat tccaagatgg 2460
tgtgaccaga catgcttcct gctccccgct tagcccacgg agtgactgtg gttgtggtgg 2520
gggggttctt aaaataactt tttagccccc gtcttcctat tttgagtttg gttcagatct 2580
taagcagctc catgcaactg tatttatttt tgatgacaag actcccatct aaagtttttc 2640
tcctgcctga tcatttcatt ggtggctgaa ggattctaga gaacctttg ttcttgcaag 2700
gaaaacaaga atccaaaacc aaaaaaaaaa aaaaa 2735

<210> 19
<211> 2822
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3256118

<400> 19
cccgtccggc cgccgccgcc gccaccgccg ccaccgcctg ggggttggtt gaggcggacg 60
gcggggtccg ggccggagta cgtcgttccc gctgcgctag gggaagcggg cagtcagaaa 120
aatgggtaag aagagtcgag taaaaactca gaaatctggc actggtgcta cagcaactgt 180
gtcaccaaag gaaatcttga acctgaccag tgagctgctg cagaaatgca gcagtccggc 240
gcctggccca ggaaaagagt gggaagagta tgtgcagatc cggactctgg ttgagaaaat 300
acggaaaaag caaaaaggtc tgtccgttac ttttgatgga aaaagagaag attactttcc 360
tgatctaatg aaatgggcct ctgaaaatgg ggcttctgtc gagggttttg aaatggttaa 420
cttcaaagaa gagggctttg gtttgagagc aacaagagat atcaaggcag aagaattgtt 480
tttatgggtt ccacgaaaat tgctaatgac tgttgaatct gctaaaaatt cagtgttggg 540
gcccttatat tctcaagacc gaatccttca agccatggga aacatcgcac tggcctttca 600
tttgctgtgt gagcgagcca gccctaactc cttctggcag ccctatattc aaaccctccc 660
cagtgaatat gacactcctc tctactttga agaagatgaa gttcggtatc ttcagtccac 720
acaagctata catgatgtct tcagccagta taaaaacaca gctcgacagt acgcctactt 780
ctataaagtc atccagaccc atcctcatgc caacaaacta cccttgaagg attctttcac 840
ttacgaggac tacaggtggg cagtctcttc tgttatgacg aggcaaaacc aaattcccac 900
agaggatggt tcccgcgtga ccctggctct gattccttta tgggatatgt gtaaccacac 960
caacggcctg atcactactg gttacaacct ggaagatgac cgctgtgagt gtgtggctct 1020
gcaggatttt cgggctggag agcagattta catttttat ggcactcgat ccaacgcaga 1080
gtttgtgatc cacagtggtt ttttctttga caataactca cacgacagag tgaaaataaa 1140
gcttggagtg agtaaaagtg acagactcta cgccatgaag gccgaggtct tggctcgtgc 1200
cggcatcccc acttccagtg tttttgcatt gcatttacc gagccgccca tctctgctca 1260
gcttttggct tttctccgag tattctgtat gactgaagaa gaactgaaag aacacttgct 1320
gggagacagc gctattgata gaatcttcac cttggggaac tcggaatttc ctgttagctg 1380
ggacaacgag gtcaaacttt ggacatttct tgaagataga gcctcacttc ttttaaaaac 1440
atataaaaca actattgagg aagataaatc cgtcttgaaa aaccacgatc tttctgttcg 1500
tgcaaaaatg gccatcaaat tgcgcttagg tgagaaagag attttggaaa aagcagtaaa 1560
gagtgcagct gtcaaccggg aatactatcg ccaacagatg gaggaaaagg ctccgcttcc 1620
caaatatgaa gagagtaacc ttgggctgtt ggagagcagc gtgggggact cgaggctccc 1680
cctggtcttg agaaacctcg aggaggaggc tggagtgcag gatgccttga acatcagaga 1740
ggcaatcagc aaagcaaagg ccacagaaaa cgggcttgta aacggtgaaa actctatccc 1800
taatgggacc aggtccgaaa atgaaagtct caatcaagaa agtaaaagag cagttgaaga 1860
cgccaaagga tcttcttcag acagcactgc tggagttaag gagtagctcg aggtgaagct 1920
ggatgggggа tccagtggag caggagttga cggacagtcc gttcacatcg ctgtgtttcc 1980
ttgttaacat ttttctttct gcagagagga agatatgttt ttgctgcttt atataaaaat 2040
ggtttttta agttatttta aaaatctagc ttccctttt gattaagatt gccatcttgc 2100
ttttaggcaa aacaaaccaa ttaacaaaca accacaagaa agggagaaga ggtgcctgtg 2160
ggagattttg cagacctatt gtgggtatag gtattttctt cctggggaag aattcagttc 2220
ccgtcttcagc tgtacttttg tgggcctgtc atcttgatga ccagaatgaa agcttgctct 2280
gcctcctgcc agccagaatt ggtggcggga cttggggata cagcgtgaag gtggggaagt 2340
tgcacagcag aaaacagaat tgaagttggg aaactctaga gtctgggcaa aatgtttggt 2400
ttttttctctt aaaaaaaata acaccccatt accaaaagaa aaggtaaggt ggcaacctta 2460
ttttttaatag tttgaaatga tgataatcct aattatataa aaatatatat ataaacacac 2520
atatatatag tgatttctaa agatttgttt acttttgtgt tttgttttac tgtactaaga 2580
acttgtcctt tctcctgaa tcaaagtagg acatgcatca tcctcctaat tttaaatgtt 2640
ggctctgatt ttaaagtggt gcatttgatt ccagccttgg taatggagag tttgcaaaca 2700
cacagcggcc cacagcttca cgtggtggtg tgcagtgtga ggcagctcct tggctttcct 2760

ggttttcaca acaagctaga gattttcaaa gctacacttt tgagtaaaaa cccttattaa 2820
aa 2822

<210> 20
<211> 1774
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4759250

<400> 20
gtgacggctg cgtgcggcgg gaatcatggc tgctcgcaga gctctgcact tcgtattcaa 60
agtgggaaac cgcttccaga cggcgcgttt ctatcgggac gtcctgggga tgaaggttct 120
gcggcatgag gaatttgaag aaggctgcaa agctgcctgt aatgggcctt atgatgggaa 180
atggagtaaa acaatggtgg gatttgggcc tgaggatgat cattttgtcg cagaactgac 240
ttacaattat ggcgtcggag actacaagct tggcaatgac tttatgggaa tcacgctcgc 300
ttctagccag gctgtcagca acgccaggaa gctggagtgg ccactgacgg aagttgcaga 360
aggtgttttt gaaaccgagg ccccgggagg atataagttc tatttgcaga atcgcagtct 420
gcctcagtca gatcctgtat taaaagtaac tctagcagtg tctgatcttc aaaagtcctt 480
gaactactgg tgtaatctac tgggaatgaa aatttatgaa aaagatgaag aaaagcaaag 540
ggctttgctg ggctatgctg ataaccagtg taagctggag ctacagggcg tcaagggtgg 600
ggtggaccat gcagcagctt ttggaagaat tgccttctct tgcccccaga aagagttgcc 660
agacttagaa gacttgatga aaagggagaa ccagaagatt ctgactcccc tggtgagcct 720
ggacacccca gggaaagcaa cagtacaggt ggtcattctg gccgaccctg acggacatga 780
aatttgcttt gtcggggatg aagcatttcg agaactttct aagatggatc cagagggaag 840
caaattgttg gatgatgcaa tggcagcaga taaaagtgac gagtggtttg ccaaacacaa 900
taaacccaaa gcttcaggtt aacggaagac atgatgcaga gcaagcctct gtgattcctg 960
cccagcacct gtgaggcctg acgtgtcagt tcccaataaa tgctcttctg atttgtttcc 1020
cgtacaggca aggaggcttg ggtagtgcag atttgtgtat ttcaatcttt gaaagctctg 1080
atgtaattta gaaatgaaat ccaatcatga gtccaggtag agaacgcctg ctgtaatcta 1140
cactgttgct gggactgcgc attctgtata taactgtgtt ggatgagtga cagatgattg 1200
tccagactag gacagcggca tgaacatgac tttggttggg attgcggata gttagggtta 1260
cctctgaatc gtgtagattt tatgagagca gctgtgcaag tgaatccaca ttaatgcctt 1320
gtcgtggtgc cattcccagc gcctgacgat acgctcttct attgtcttat tctggcaggt 1380
tttgacgttt taaatttttt aaagaaattt tattccttgg accaaaaggt ttggttaacc 1440
acccccctct tacttgcttt cacattttga gtgtccagag gaaacagaaa ggaatgagtg 1500
tgtgacgttg ctgcacgcct gactctgtgc gagcttcttt ctgtgtatat attttgtttt 1560
attttttcc gtgtatattt ttaatcccga cagaacatca tgtgagattt ctttaaaatg 1620
gattaaacga tttctttcagc ctgaaaaaaa aggttttgaa aatgttttct tgtagttttg 1680
tttggttcta aacaacaaat aggttttaat cactcgaaat ggaattatat tgtgtattca 1740
ttgaataaat ttttttttgaa agtaaaaaaa aaaa 1774